TISSUE ANTIGENS
IMMUNE RESPONSE GENETICS

Tissue Antigens ISSN 0001-2815

Genetic differentiation of Jewish populations

W. Klitz[1,2], L. Gragert[3], M. Maiers[3], M. Fernandez-Viña[4], Y. Ben-Naeh[5], G. Benedek[6,7], C. Brautbar[6,7] & S. Israel[7]

1 School of Public Health, University of California, Berkeley, CA, USA
2 Public Health Institute, Oakland, CA, USA
3 National Marrow Donor Program, Minneapolis, MN, USA
4 Department of Laboratory Medicine, University of Texas, MD Anderson Cancer Center, Houston, TX, USA
5 Department of History of the Jewish People, Hebrew University, Jerusalem, Israel
6 The Lautenberg Center for General and Tumor Immunology, Hebrew University – Hadassah Medical School, Jerusalem, Israel
7 Tissue Typing Unit, Hadassah University Hospital, Jerusalem, Israel

Key words
Jewish; HLA haplotypes; population differentiation

Correspondence
William Klitz
50 University Hall
School of Public Health
University of California
Berkeley
CA 94720-7360
USA
Tel: +1 510 643 1594
e-mail: klitz@berkeley.edu

Received 22 February 2010; revised 7 May 2010, 8 July 2010; accepted 13 July 2010

doi: 10.1111/j.1399-0039.2010.01549.x

Abstract

The Jewish diaspora can be viewed as a natural process in population dispersion and differentiation. We extend genetic studies on the Jewish diaspora to an analysis of human leukocyte antigen (HLA) haplotype distributions in the Jewish peoples, and show the value of this information for the design of Jewish marrow donor registries. HLA data from the Hadassah Bone Marrow Registry having parental country-of-origin information comprise samples of geographically discrete regions. We analyzed the HLA allele and haplotype frequencies for each national sample using population genetic and clustering methods. Population differentiation among diaspora populations was shown on the basis of HLA haplotype frequencies, including differences within the more recently diverged European groups. A method of haplotype and population clustering showed patterns of unique haplotype affinities associated with specific Jewish populations. The evidence showed that diaspora Jewish populations can be sorted into distinct clades of which the Ashkenazi are but one. Relationships among Jewish populations are interpretable in light of the historical record. We suggest that a major contributing factor to the genetic divergence between Jewish groups may have been admixture with local host populations, while, at the same time, threads of Eastern Mediterranean ancestry remain evident.

Introduction

This work adds to the growing knowledge of genetic variation in humans by showing the power of a single gene complex, human leukocyte antigen (HLA), for tracing the population history of the Jewish peoples. The major histocompatibility complex (MHC) of vertebrates – the HLA system of humans – is characterized by remarkable levels of population differentiation standing alongside its high genetic diversity. These qualities make it especially valuable as a tool for the study of population divergence, natural selection and evolution in humans (1). The function of HLA variability is central to a host's relationship to microbial agents – through increased capacity for pathogen pattern recognition genetic heterozygosity raises the likelihood of individual survival. Here, we analyze HLA diversity using genetic and population data from an Israeli bone marrow registry providing novel information on the genetic differentiation of Jewish peoples worldwide.

In a study of population expansion and diversification in an organism, genetic markers compared at the population level supply a useful basis for understanding evolutionary modification. From this strictly biological basis inferences of change or stasis can be sought and interpreted. In the case of human populations which have migrated and diversified more recently, the historical record itself can prove a valuable adjunct for comparison to the genetic record. Through Jewish history two conflicting forces have operated on diaspora populations – one maintaining population integrity and resisting admixture with outsiders, and the other leading to the integration of other non-Jewish backgrounds into the Jewish gene pool. These divergent tendencies can be illustrated with two examples from Jewish antiquity. The Bible describes Mosaic Law discouraging contact with gentiles, including the establishment of restrictions on intermarriage. The contrast is the early appearance of rabbinical specifications for the conversion of gentiles to Judaism (2).

In biological terms, Judaism contains cultural prescriptions promoting genetic privacy, while at the same time specifying conditions allowing exogamy. These cultural circumstances set the stage for a study of the genetic outcomes of one of the best recorded population expansion and diversification events in our species – that of the Jewish diaspora.

Cultural forces of the last century have resulted in many members of the Jewish diaspora moving to Ottoman and British-mandated Palestine, which later became the modern state of Israel. Here, we use the population records and HLA genotyping data available from an Israeli bone marrow registry to examine differences in the Jewish people at the population level. The results are compared with previous studies of Jewish genetics, to HLA-specific studies, and to the historical record.

Historical context of influences on the genetic composition of Jewish people

A brief overview of Jewish population movements in history (3–5) sets the stage for the interpretation of genetic differences in the context of their modern manifestations. A founding narrative of the early Jewish nation claims that the people of Israel are a direct descendant of Abraham, Isaac, Jacob and his sons. The historical record, however, shows that the Jewish people were not drawn on such a narrow genetic base. Early exile experiences can be traced back to the fall of the Kingdom of Israel to the Assyrians in 712 BCE that gained force with the expulsion of many of the inhabitants of Judea to Babylonia (territories which are now in Iran and Iraq) by the Babylonian conquerors immediately after the destruction of the First Temple in Jerusalem (586 BCE). These early emigrants constituted some of the first Jewish communities in exile, some of whom persisted continuously through to the 20th century.

During the Hellenistic period occurring in the BCE centuries between the Babylonian era and the Roman period, Jewish settlements grew in Egypt and Syria adjacent to Palestine. In the Roman period, beginning in the 1st century BCE, significant Jewish communities could also be found in Western Anatolia (modern Turkey), Greece and Italy, and at least some individual presence in Northern Africa, in Spain and within the European continent. During the following centuries, the map of Jewish settlement expanded east to the Byzantine Empire and north to the Balkans. From Italy, Jews spread north to the important trade centers in Central and Western Europe, and from Egypt, Palestine and Abbasid (Babylon from 750 to 1258 CE) they migrated westward and settled along the North African coast. In the East, Jews reached far into Asia, beyond the Persian speaking countries, as far as India and China.

In addition to immigration and natural increase, another important demographic factor was the conversion of individuals and groups to Judaism. There is evidence of a conversion movement within Syria and in the Greek–Roman cities in the Roman Empire, including Rome itself, with some of these converts becoming fully integrated into existing Jewish communities. At various times, forced conversions also enlarged the Jewish genetic base, exemplified by conversions of the conquered Edomites and Itureans, north and south of Jerusalem by the Jewish rulers of the late 2nd century BCE. Another source of genetic input was through the conversion of slaves, as is documented from the Ottoman Empire (6).

Jewish cultural forces operated throughout the diaspora to maintain Jewish coherence. Alongside various folk traditions (7), Rabbinical Judaism had a major role in the maintenance of long lasting cultural uniformity independent of any single geographic focus. These behaviors added immeasurably to the survival of Jewish people by consistent recognition of common cultural roots and maintenance of a shared heritage and religious practices.

Population divergence at the HLA complex

Numerous publications have shown the great informativeness of HLA haplotypes from population studies (8). In this study, we capitalized on population divergences created over a period of some 80 generations deriving from individuals who had emigrated from the Eastern Mediterranean to establish new populations throughout much of the known world. A reversal of these population movements has since occurred over the last century as many Jews migrated to the modern state of Israel. These events made possible collections on which this work is based and permit an examination of the degree of differentiation at the HLA complex among populations of this most famous diaspora. The historical documentation of many of the details of population dispersal and subsequent demographic events also help to clarify the results of our genetic analysis. In addition, from a biological viewpoint, these circumstances create a unique opportunity to examine evolutionary change in human populations derived from a single geographic source.

Previous genetic studies of Jewish populations

Identification of remnant threads of Jewish populations based on data from current members of the Diaspora has yielded some surprising genetic data, especially from Y chromosome haplotypes, in which Cohen priestly castes were found to show instances of remarkable coherence. The most astounding instance of these studies is shown by the Lemba people in Southern Africa, who have been found to carry the Cohen Y chromosome haplotypes among their male members (9). The Cohen Y chromosome has since been traced to many populations across the Jewish diaspora (10). Other Y chromosome studies have been traced through another class of priests, the Levites (11). Mitochondrial DNA (MtDNA) variation has also identified particular demographic characteristics of the Jewish expansion (12, 13). These studies show that many Jewish populations are derived from peoples who inhabited the eastern Mediterranean region.

Previous studies have also shown the genetic introgression of Jews into their host populations. A study of Y chromosome variation on the Iberian Peninsula showed the presence of Jewish-characteristic Y haplotypes in contemporary Iberian males (14). Jews resided in Iberia for centuries up to the expulsions and forced conversions of the late 14th and 15th centuries. The Jewish contribution to the contemporary Iberian male population was found to be 20% averaged across provinces of Spain and Portugal. In additional examples, an HLA signature distinctive of Jews has been seen in Mexican American samples from the United States (15), and similarly in samples of US Caucasians (Martin Maiers, personal communication).

The discriminatory ability of genome-wide surveys of Jewish populations is shown through studies on Ashkenazi Jews, in which a set of loci selected from a large genetic screen is capable of identifying Ashkenazi presence within European populations with great confidence (16). This work has been extended to examine the degree of Ashkenazi and European ancestry, showing that complete discrimination of Ashkenazi with only one Jewish grandparent from that of Europeans is possible (17, 18).

The HLA complex has been studied in Jewish populations for over 30 years. Even in early studies some distinctive characteristics of Jewish population samples were suggested using the serological reagents available for HLA typing at the time. Some of our current observations were reflected in these earlier studies. Precursors to the A26 and B38 antigens (i.e. A10 and B16) were found to be frequent and in disequilibrium (19). A comparison of HLA-A and -B antigen frequencies showed similarities between a sample of Armenians and Iranian Jews, but the haplotypes tended to be distinct (20). The Cochin Jews of India carry the common Jewish allele A26, but without B38 (21). In a study of HLA and other genetic markers among non-Ashkenazi, Cochins were found to be most distant (22). It was then realized that the many populations of non-Ashkenazi Jews constitute a heterogeneous assemblage (23). The application of molecular HLA typing methods uncovered additional distinctive characteristics, including the identification of a common DRB1*1305 allele in Moroccan Jews (24), and the DR4 component of the A*26-B*38 haplotype was identified as DRB1*0402 in Ashkenazi Jews (25, 26). The sharing of A*26-B*38-DRB1*04 and other distinct haplotypes among the Ashkenazi and the Moroccan Jews was clarified (27). Ethiopian Jews (too rare in our state-of-origin identified classification from the Hadassah Registry for inclusion) were found to be outliers among Jewish populations, with none of the classic indicators of Middle East origins. For example, A*26 is common but B*38 completely absent. At the same time, the Ethiopian Jews, while showing some HLA affinities with samples of native Ethiopians, retain distinctive HLA characteristics (28) (Chaim Brautbar personal communication). Our current study overcomes the limitation of previous work by including much larger sample sizes and more accurate and precise HLA genotyping methods. A final important departure of this current work from previous studies is the simultaneous assessment of HLA variation in many Jewish populations.

Methods

Population resources

This analysis uses all 55,801 DNA-typed donors from the Hadassah registry in Israel. At recruitment, the donors were asked their country of origin and the country of origin of each parent. A summary of country-of-origin information for cases in which both parents were from a single country and $n > 50$ is listed in Table 1. Many other country-of-origin groups with smaller sample sizes were present in the original extraction, along with individuals of mixed country-of-origin provenance. These were all excluded from the analyses.

One population referred to as 'Combined' consists of individuals whose ancestral origin is unknown, mixed or native to Israel. The large size ($n = 45{,}509$) and predominantly Ashkenazi origin of this group make it valuable for validation of the various methods. The West African, European and Native American haplotype frequency distributions were used as out-group samples for constructing a dendrogram of Jewish population relationships. The West African and Native American frequency distributions were derived through a method of subtracting haplotype frequencies of a parental population from an admixed group (29) using National Marrow Donor Program population samples of African Americans, European Americans and Mexican Americans.

HLA typing

All DNA-typed volunteer donors were included in the analysis. A total of 23,856 were typed for HLA-A, -B and -DRB1 loci. The remaining 31,945 donors for whom only HLA-A and HLA-B loci were genotyped were also included, using an unbiased estimation procedure to obtain three-locus haplotypes and allele frequencies (30). The use of this large, unselected sample of HLA-A and -B typings was useful in reducing sampling bias observed in registry samples typed at several loci, including those high-resolution typed. Altogether 7604 individuals were available who had been typed for A, B and DRB1 at high resolution. Donors with consistent country-of-origin information totaled 12,752. Serologically typed donors were excluded because of the increased error rate associated with this typing method.

HLA typing results at the intermediate or high-resolution level were collapsed into a hierarchical two-digit HLA nomenclature for this analysis (31) (see also the National Marrow Donor Program website). In cases where subdivisions of broad types were available in sufficient quantity, the estimation was made at the split level (e.g. DR05 estimated to DR11 and

Table 1 Jewish population samples and their use in the various analyses

Region/Country	2*n*	Populations used in each analysis: Common haps – Table 2; Hap frequencies – Table S1; Hap correlations – Figure 2	Signif. Table 3	Signif. Figure 1A	Signif. Figure 1B	NJ tree[a] Figure 3	PCA Figure 4	CLUTO Figure 5
North Africa								
Algeria	184	x		x		x		
Egypt	320	x		x		x	x	x
Libya	428	x		x		x	x	x
Morocco	3768	x		x		x	x	x
Tunisia	670	x		x		x	x	x
Western Asia								
Afghanistan	102	x		x				
Georgia	254	x		x		x	x	x
Iran	1180	x		x		x	x	x
Iraq	1770	x		x		x	x	x
Turkey	512	x		x		x	x	x
Yemen	1030	x		x		x	x	x
India[b]	750	x		x		x	x	x
Europe								
Bulgaria	166	x			x	x		
Czech	150	x			x			
France	530	x			x	x	x	x
Germany	396	x			x	x	x	x
Hungary	288	x			x	x	x	x
Lithuania	110	x			x			
Netherlands	132	x			x			
Poland	1676	x	x	x	x	x	x	x
Romania	1852	x			x	x	x	x
Russia	2566	x	x		x	x	x	x
Switzerland	102	x			x			
Ukraine	610	x			x	x	x	x
UK	648	x			x	x	x	x
Emigrant Europe[b]								
Argentina	900	x				x	x	x
Brazil	146	x						
Canada	262	x				x	x	x
South Africa	362	x				x	x	x
Uruguay	152	x						
USA	3136	x				x	x	x
Combined[b]	85,018	x				x		

[a]Also uses the out-groups (2*n*) West African (165,412), Native American (12,664) and European (824,292).
[b]Populations derived from emigrations of European Jews.

DR12). When few results were available at the split level, all types were rolled to a broad level (e.g. DR15 and DR16 were rolled up to DR02). The rule was that if the polymorphism of the rarer allele of a broad allele split was greater than 20%, then the splits were used. The sample sizes and the percentage of samples characterized at high resolution were too small to give accurate four-digit HLA frequencies per population. However, using high-resolution subsets of the data, it was possible to infer the most likely high-resolution HLA linkages within the two-digit haplotypes of the most common alleles. Three-locus HLA-A, HLA-B and HLA-DRB1 haplotypes are indicated in a shorthand with separating tildes, e.g. 26∼38∼04.

Population genetics analyses

Haplotype frequency analyses

To estimate haplotype frequencies from unphased phenotypes, we used an implementation of the expectation-maximization (EM) algorithm that handles mixed resolution HLA typing and missing DR locus data (30). Frequency distributions for each country/region were estimated separately. While this study did not have enough high-resolution typing data to ascertain four-digit A∼B∼DRB1 haplotype frequencies for each country independently, using the pooling high-resolution data from all country-of-origin categories (described above) and estimating four-digit frequencies allowed assignment of

the most common two-digit haplotypes to their most likely four-digit subtypes. These high-resolution estimations were used in the figures and tables of the haplotype results, and are made available at http://bioinformatics.nmdp.org/Jewish. The resulting haplotype frequencies are estimations and thus inexact, although it has been shown that estimation variance is modest (32). In our work here, the fact that allele frequencies alone show the same tendencies to population differentiation as haplotype frequencies suggests that estimation variance in haplotypes cannot be strongly influencing the inferences drawn from these values.

Exact tests for Hardy–Weinberg equilibrium (HWE) were run for each locus (33). While HWE is assumed for haplotype estimation, only limited deviations accrue in the face of small HWE departures (34).

Genetic distance

To visualize genetic distances between the Jewish populations, we used a set of tools from the PHYLIP program to make a population dendrogram. First, the Seqboot module created 1000 resampled versions of the input dataset of haplotype frequencies using the bootstrap shuffling method. Matrices of Nei's genetic distance were then calculated using the Gendist module for comparing populations (35). A total of 1000 dendrogram replicates based on the matrices of Nei distances were generated using the nearest-neighbor algorithm in the Neighbor module. The final consensus dendrogram combined these trees using the Consense module and gave the bootstrap supports for each subtree (36).

Principal components analysis

Principal components analysis (PCA) was performed on the entire haplotype frequency distribution of each population using the princomp function in MATLAB. 2-D PCA plots were created using the pairs of principal components to display the haplotypic variation among populations. PCA extracts independent vectors of explained variance, beginning with PC1 which explains the most variance. We take advantage of the fact that even principal components explaining smaller fractions of total variance may be interpretable with population data (37).

Clustering analysis

HLA haplotype frequency visualizations on Jewish populations were created using the software package clustering toolkit (CLUTO) (38 and http://www-users.cs.umn.edu/~karypis/cluto/). The first 100 frequency-ranked haplotypes were put into a neighbor-joining (NJ) tree using the cosine algorithm based on their frequency pattern across countries. The number of clusters is specified and is based on partitioning the tree on the distance from the root. A second independent step rearranges the countries based on a neighbor-joining tree using the pattern of frequencies across all the haplotypes. The tree is drawn with the default option (hierarchical agglomerative clustering). The final matrix results in rearrangement of haplotypes and populations to maximize block density to illustrate the population privacy of many of the HLA haplotypes. Within and among cluster differences are expressed with ISIM and ESIM (Internal and External SIMilarity) statistics, respectively. The CLUTO program performs bi-clustering to cluster haplotypes and populations based on haplotype frequencies in the same way that genes and subjects are clustered in gene expression array profile analysis. It should be noted that the sum correlation, PCA and NJ analyses each operate on the complete A~B~DRB1 haplotype frequency distributions of the populations, while CLUTO examines only the top 100 haplotypes. For this reason, the CLUTO results may express more dominant themes of a population's composition, and miss, e.g. admixture events of lesser impact.

Results

In the initial foray into Jewish population genetic contrasts we present a general overview of the HLA single locus variation among populations, followed by the frequencies of the top 100 A~B~DRB1 haplotypes found in Jews on which much of the further analysis is based. From the outset it is not known whether supportable statistical differences are present between the Jewish population samples. This is assessed using a test of homogeneity between pairs of populations. Relationships among populations are then addressed and displayed with a series of methods including paired correlations of haplotype frequencies, a neighbor-joining tree, PCA and a two-way (haplotype by population) clustering tool. A final result stemming from the previous analyses presents lists of Jewish haplotypes attributable to the Eastern Mediterranean region from which the Jewish diasporas are understood to have originated.

The operational units of study in this work are the national categories defining an individual's parental origin. Table 1 presents sample sizes for the country-of-origin populations grouped by regional affinity. The Ashkenazi can be thought of as an ethnicity within Jews – here tentatively represented by those individuals originating in Central and Eastern Europe including samples from Czech, Germany, Hungary, Lithuania, Poland, Romania, Russia and the Ukraine, as well as emigrant populations from these regions. An overview of the populations used in each data presentation or analysis – influenced by variation in sample sizes among the populations – is also shown in Table 1. All 32 population samples, including the composite sample 'Combined', are used in presenting most common haplotypes (Table 2), haplotype frequencies (Table S1) and population correlations (Figure 2). The three analyses of statistical differences use three sets of populations – including two tests of closely related European groups (Table 3), primary Jewish population lineages having 12

Table 2 The five most common HLA A∼B∼DRB1 haplotypes for 32 Jewish populations grouped by region and country of origin

		Top five-ranked A∼B∼DR haplotypes				
Region/Country	Sum top 5	H1	H2	H3	H4	H5
North Africa						
Algeria	0.1843	01∼44∼06	03∼13∼07	26∼38∼04	02∼52∼11	24∼18∼11
Egypt	0.1518	24∼35∼04	24∼50∼07	01∼35∼11	03∼35∼11	33∼14∼01
Libya	0.2347	02∼50∼07	24∼35∼11	01∼35∼07	31∼40∼04	02∼41∼06
Morocco	0.1491	03∼13∼07	01∼13∼07	24∼18∼11	26∼38∼06	26∼38∼04
Tunisia	0.1722	11∼35∼06	23∼44∼11	02∼50∼07	01∼35∼11	02∼44∼11
Western Asia						
Afghanistan	0.1255	01∼35∼11	26∼38∼06	23∼44∼11	01∼50∼07	03∼18∼11
Georgia	0.2542	03∼35∼11	68∼15∼06	23∼49∼11	03∼35∼06	01∼15∼06
Iran	0.1898	24∼18∼11	01∼52∼06	11∼53∼06	30∼53∼03	01∼15∼06
raq	0.1835	03∼44∼04	01∼15∼06	03∼35∼11	01∼41∼07	01∼35∼11
Turkey	0.1145	01∼35∼11	02∼18∼11	68∼14∼01	24∼38∼06	24∼35∼11
Yemen	0.1918	23∼44∼07	33∼14∼12	02∼51∼04	69∼35∼04	01∼44∼06
India[a]	0.1966	33∼44∼07	29∼07∼10	24∼35∼11	03∼44∼04	24∼57∼04
Europe						
Bulgaria	0.3502	68∼14∼01	26∼38∼04	33∼14∼01	24∼18∼11	30∼35∼06
Czech	0.2050	01∼57∼06	26∼38∼04	24∼35∼11	01∼35∼11	02∼14∼06
France	0.1438	26∼38∼04	01∼35∼11	24∼35∼11	68∼14∼01	30∼13∼07
Germany	0.1932	26∼38∼04	01∼35∼11	01∼57∼06	24∼35∼11	26∼38∼06
Hungary	0.2363	26∼38∼04	24∼35∼11	33∼14∼01	01∼35∼11	02∼35∼12
Lithuania	0.2428	26∼38∼04	11∼52∼02	01∼57∼06	01∼35∼11	02∼15∼06
Netherlands	0.1741	02∼35∼11	26∼38∼04	24∼35∼11	01∼08∼09	02∼35∼01
Poland	0.1746	26∼38∼04	01∼57∼06	24∼35∼11	68∼14∼01	33∼14∼01
Romania	0.1862	26∼38∼04	24∼35∼11	33∼14∼01	01∼57∼06	01∼35∼11
Russia	0.1455	26∼38∼04	24∼35∼11	01∼57∼06	33∼14∼01	01∼35∼11
Switzerland	0.2020	03∼07∼07	02∼35∼04	26∼08∼12	24∼35∼04	26∼38∼04
Ukraine	0.1778	26∼38∼04	24∼35∼11	33∼14∼01	01∼35∼11	02∼44∼04
UK	0.1639	26∼38∼04	24∼35∼11	01∼57∼06	33∼14∼01	11∼52∼06
Emigrant Europe						
Argentina	0.1365	26∼38∼04	24∼35∼11	01∼35∼11	33∼14∼01	68∼14∼01
Brazil	0.2594	26∼38∼04	24∼35∼11	24∼38∼11	01∼57∼06	24∼14∼01
Canada	0.2219	26∼38∼04	02∼35∼11	01∼57∼06	33∼14∼01	24∼35∼11
South Africa	0.1779	26∼38∼04	01∼57∼06	01∼35∼11	11∼52∼06	03∼38∼06
Uruguay	0.1978	26∼38∼04	03∼14∼01	01∼08∼03	02∼35∼11	24∼35∼11
USA	0.1596	26∼38∼04	24∼35∼11	01∼35∼11	01∼57∼06	33∼14∼01
Combined[a]	0.1239	26∼38∼04	24∼35∼11	01∼35∼11	01∼57∼06	33∼14∼01

[a]Significant single locus Hardy–Weinberg deviations after adjustment for number of tests – India (HLA-A) and Combined (HLA-A and -DRB1).

Table 3 Statistical differences between HLA allele and haplotype frequencies for two Ashkenazi Jewish populations, Russia and Poland

	A	B	DR	A∼B∼DR
k	16	25	12	186
G	63.9	157.8	152.8	1435.8
P	1.1E-07	3.1E-21	1.5E-26	8.8E-192

k, The number of alleles or haplotypes tested, G, the log likelihood ratio statistic; *P*, the *P* value.

non-European Jewish populations as well one representative European group (Figure 1A), and the Europe-only comparisons using 13 populations (Figure 1B). The genetic distance and clustering methods use only subsets of the populations in order to reduce possible spurious results because of smaller sample sizes. The PCA (Figure 4) and CLUTO (Figure 5) analyses use only samples with $2n > 200$, while the neighbor-joining analysis (Figure 3) relaxes that limit somewhat in order to include the Algerian and Bulgarian samples.

Common alleles and haplotypes

Additional diversity underlies most of the alleles defined to two-digits, but even at two-digit resolution it is evident that all major HLA allelic lineages are present. At high resolution, a number of HLA alleles relatively common in Jewish populations and the Ashkenazi in particular stand out as being rare or absent in Europe (presented and discussed further below). Among these are A*0205, A*0302, B*0705, and the six DRB1 alleles *0102, *0402, *1104, *1303, *1305 and *1502. At low (two-digit) resolution these distinctive population signatures may be missed.

We outline the single locus HLA frequencies across the 31 country-of-origin populations and the 'Combined' group, supplying an initial overview of HLA variation. Some differences in allele and haplotype diversity may be attributable to sample size differences among groups. The polymorphism of the HLA-A locus in country-based samples ranges from 14 alleles (Algeria) to 20 (Morocco). As an indication of the high within population diversity characteristic of HLA loci, the heterozygosity ranges from 87% in Lithuania to 90% in Argentina. The allele A*02, common and often most frequent in human populations worldwide (8), is the most common allele in all but the small sample of Czech Jews. This same pattern is also true for the 12 non-European Jewish populations with three exceptions – in Algeria, Iran and India A*02 is the fourth or fifth most common A allele with frequencies of 12.5%, 10.5% and 12.5%, respectively. Although relatively uncommon in most human populations except for Southeast and Northeast Asia (8), A*26 is second or third most common in all but 5 of the 17 Ashkenazi or 'Emigrant Europe' Ashkenazi populations, and fourth most common in those five cases. None of the 12 other Jewish populations bear A*26 in the most frequent alleles. Although A*26 is less common in the 'Other Jewish' super-group in which the frequency is 4%–8%, in Libya and Yemen the frequency drops to 2.3 and 1.9, respectively. Alleles A*01, A*24 and A*03 are also common across most of the Jewish population samples. A*23 is uncommon in most regions worldwide except sub-Saharan Africa where it exceeds 20% and in Europe and North Africa where it is near 8% (8). While A*23 is found at less than 5% frequency across the Ashkenazi samples and in most of the other Jewish groups, the frequency rises to 14.1% in the Yemeni Jewish sample, where it is the third most common A allele.

The most polymorphic major histocompatibility locus, HLA-B has from 20 to 30 alleles in each country-of-origin sample. Heterozygosity at HLA-B ranges from 86% in Georgia to 93% in India. B*35 is a most common allele in all except four populations (Algeria, Morocco, Yemen and Czech) where it is still common in each instance. The Ashkenazi, both Europe and 'Emigrant Europe' groups, are remarkably consistent with nearly all populations having the three most common B alleles in the order B*35, B*38 and B*14.

Two-digit typing resolution at the DRB1 locus yields 14 alleles across all 32 population samples. Population heterozygosity at DRB1 ranges from 80% in Iraq to 89% in the Netherland Jewish sample. The DRB1 alleles, *11, *04 and *07 are the three most common alleles in nearly all populations. DRB1*01 replaces DRB1*07 as third most common in six of the Ashkenazi populations, and the DRB1 alleles *01, *03, *10, *13 and *14 appear as second or third most common in several of the non-Ashkenazi populations.

Genotypic proportions of HLA-A, -B and -DR showed that most populations and loci were found to be in HWE. Before adjustment for multiple testing, the 32 populations deviated in HWE at $P < 0.05$ in three HLA-A tests, four HLA-B and five HLA-DR tests, when 1.6 such instances would be expected at each locus by chance. After adjusting for the 32 tests carried out at each locus, only the DRB1 samples for India and at HLA-A and -DRB1 'Combined' significantly deviated from expected genotypic ratios. Each of these instances might be anticipated because of the presence of distinct subpopulations: 'Combined' is a composite of all Israeli Jewish populations, and the India sample is comprised of both the distinctive Cochins and other India-origin Jews. These deviations are explained by the Wahlund effect because of population substructure within a single sample. For the group 'Combined' this gains support from the excess of observed homozygotes in each case: 15.4% *vs* 14.3% expected for HLA-A and 20.1% *vs* 19.2% for HLA DRB1, and similarly, for the DRB1 locus in India in which the observed and expected fractions of homozygotes were 22.9% and 14.7%. As a whole, the various population samples fit Hardy–Weinberg genotype proportions with the few exceptions explainable by recognized population substructure.

The sums of the frequencies of the most frequent five A∼B∼DRB1 haplotypes (designated H1 to H5) range from 11% in Turkey to 26% in Brazil (Table 2). This constitutes a minority of the haplotype information, but still can yield a useful overview of haplotype variation. Individuals in the large 'Combined', although of mixed provenance, are predominantly of Ashkenazi origin, and so render a defensible snapshot of the most common HLA haplotypes in the Ashkenazi. Of these, the haplotype 26∼38∼04 is present in all Europe and Emigrant Europe groups and usually at rank H1. We note that Bulgaria (on the Europe–West Asia border) at H2 and Morocco at H5 also carry this as a common haplotype. The second most common haplotype, H2 in the 'Combined' group, 24∼35∼11, is also in the top five haplotypes of all 18 European-origin samples except for Lithuania, Switzerland and South Africa. For Libya and Turkey this is also among the most common haplotypes. The remaining three haplotypes in the 'Combined' top five are also well represented in the European-origin groups and occasionally in the non-European populations. For example, 01∼35∼11 is prominent in five non-European groups and of first rank in Afghanistan and Turkey. The H5 haplotype in 'Combined', 33∼14∼01 is also at H5 in Egypt and at H3 in Bulgaria.

Taken as a whole, the European Jewish populations show considerable haplotype homogeneity for the most common haplotypes. Switzerland and the Netherlands are possible outliers in this regard, as evidenced by sharing only two and one haplotypes with the 'Combined' top five, although the small sample sizes of each must be taken into account. The 12 non-European Jewish samples show only limited or a complete absence of sharing of the most common Ashkenazi haplotypes (as defined by the largely Ashkenazi 'Combined' sample). The origin and interpretation of these many distinctive and largely unshared haplotypes are evidence of historic demographic events experience by these groups. In this regard, we note

that two common Northern European haplotypes 01∼08∼03 and 02∼44∼04 are seen in the European-origin samples of Uruguay and the Ukraine, respectively.

The 100 most common A∼B∼DRB1 haplotypes

Three-locus A∼B∼DR haplotype frequencies were estimated for each country-specified group separately. Table S1 summarizes the haplotype frequencies by rank of frequency for the first 100 haplotypes across 31 country-of-origin defined populations and 'Combined'. Haplotypes were ordered by the most common haplotype in any population. For example, the first haplotype, 23∼44∼07, found in Yemen at 10.0%, was the most common haplotype in any single population. It was rarely observed or absent in other populations. The second-ranked haplotype 26∼38∼04 was the next most commonly observed (at 9.6% in Lithuania), but also found to be quite common across a range of other populations. The 100th most common haplotype (02∼14∼01) was present in Brazil at 1.96%. There were no haplotypes below this frequency in any of the Table S1 populations. The 100 haplotypes comprised from 32.1% (Yemen) to 59.3% (Brazil) of the total sample within populations. The seven populations for which haplotype sample sizes were close to or less than 150 are indicated with hatching (Table S1). These smaller samples can be unstable in estimates of haplotype frequencies, but are presented for the sake of completeness and because they bolster general observations regarding Jewish populations as a whole.

Using the 7604 individuals from the Hadassah Registry who were typed at four-digits (high) resolution at A, B and DRB1, we estimated the high-resolution composition of the 100 frequency-ranked haplotypes (Table S1). We believe that the great majority of these high resolution haplotypes are likely to be true and accurately estimated, nonetheless, their creation through an estimation procedure should be kept in mind. A list of all high-resolution typed HLA haplotype and allele frequencies are available at the National Marrow Donor Program website http://bioinformatics.nmdp.org/Jewish/.

Statistical testing of population differences by HLA allele and haplotype frequencies

The statistical magnitude of differences between the Jewish population samples was shown using pairwise tests for homogeneity. Significant differences between populations were found to be nearly universal. This is exemplified by the difference between two Ashkenazi samples (Russia and Poland), expected to differ very little. We display the results for this particular pair of population comparisons because of their relatively robust samples sizes (Table 3). The frequencies of the 16 HLA-A alleles tested for the Russian and Polish comparison resulted in a log likelihood ratio statistic of 63.9 with 15 degrees of freedom, giving $P = 1.1E\text{-}07$. The HLA-B and HLA-DRB1 tests show even greater significance between the two populations. The A∼B∼DR haplotype comparison showed extreme differences, with $P = 8.8E\text{-}192$.

Significant differences in the Jewish population lineages as a whole are presented by plotting the distribution of P-values for each pair of populations. In this comparison, Poland was used as a single Jewish population representative of European Jews, along with the 12 non-European samples. The P-values for the pairwise differences among the 12 Jewish country-of-origin populations for HLA-A and for the A∼B∼DR haplotypes are shown in Figure 1A. The distributions of single locus HLA-B and -DRB1 tested P-values were similar to that described for HLA-A. Only one comparison of HLA-A allele frequencies failed to reach nominal significance, while 76% (69/91) of the comparisons of HLA-A differed at $P < 0.001$. For A∼B∼DR haplotypes the distribution of P-values was shifted to the right, with 99% (90/91) of the comparisons exceeding $P < 0.001$.

Restricting the analysis to the 12 European samples, which include a number of different Ashkenazi groups (and excluding the six populations with $2n < 160$ and 'Combined'), differences were, as expected, much smaller compared with that of the Jewish diaspora as a whole. Nonetheless many of the European populations differed significantly (Figure 1B). A few of the HLA-A locus comparisons did not reach nominal significance, especially after applying multiple test adjustments to the alpha value. The distribution of values for the A∼B∼DR haplotype comparisons was again shifted to the right and all but one of the comparisons (Hungary *vs* 'Combined' with $P = 3.16E\text{-}02$) had high P values. In summary, HLA single loci and haplotypes showed formally statistically significant differences among populations, even among the more recently diverged and mostly Ashkenazi European Jewish populations.

Summed correlations

These same tendencies (showing affinities among the European Jews and dissimilarities among the remaining populations) were evident in the summed pairwise Spearman rank correlations of haplotype frequencies that measured the similarity of each population against all others. This is illustrated by a depiction of pairwise correlations among haplotype frequencies (Figure 2). The US and Combined samples top the list with greatest similarity to all other populations taken as a whole. The European samples tend to have the highest correlations (most cluster in the bottom right portions of the figure) because of the large number of more similar Ashkenazi populations.

Population tree

The results of a neighbor-joining analysis using Nei genetic distances among 27 Jewish population samples and three outgroups (West Africa, Native America and Europe) using the three-locus HLA haplotype frequencies are shown in Figure 3.

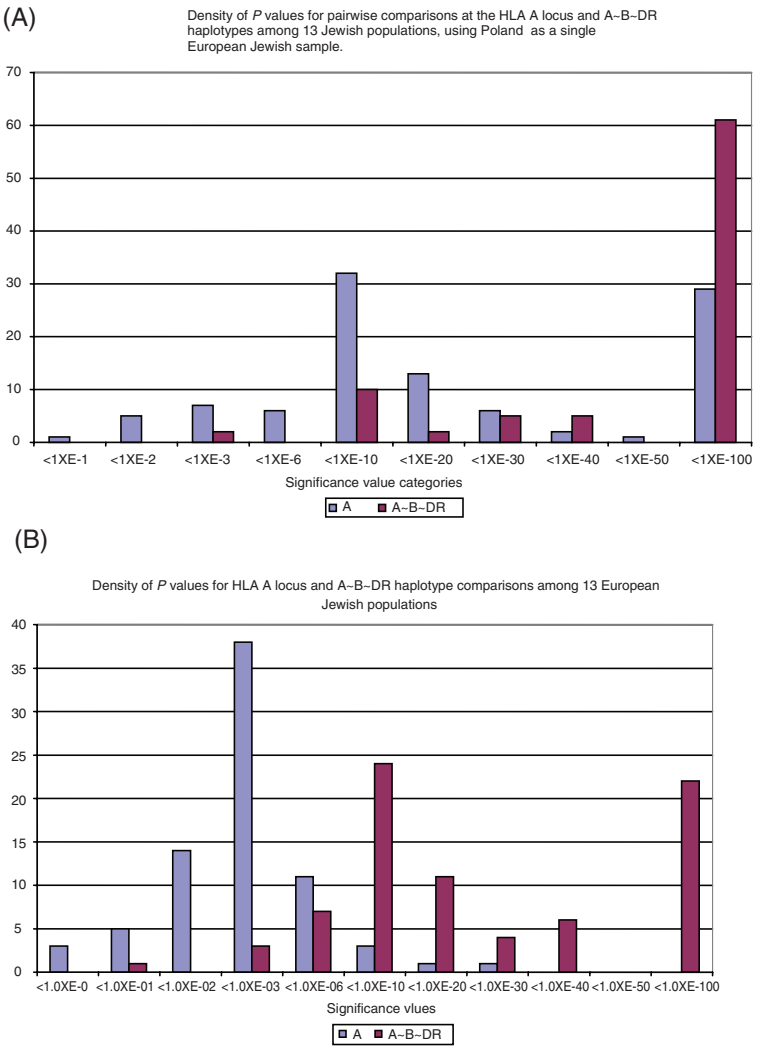

Figure 1 (A) Density distribution of *P* values for HLA-A locus and A∼B∼DR haplotypes for the 91 pairwise comparisons of 13 Jewish populations. The countries include Afghanistan, Algeria, Bulgaria, Egypt, Georgia, India, Iran, Iraq, Libya, Morocco, Poland (as representative European), Tunisia, Turkey and Yemen. (B) Distribution of *P* values for the 78 pairwise comparisons of 12 European Jewish populations and the 'Combined' group. Countries include Czech, France, Germany, Hungary, Lithuania, Netherlands, Poland, Romania, Russia, Switzerland, Ukraine and UK. Class limits for each category of significance are indicated with the less extreme limit. The more extreme limit is the next category to the right. For example, the category indicated '<1XE-06' has class limits of 1XE-06 to <1XE-10.

The cluster of populations extending from Russia to Ukraine (Figure 3) encompasses the Ashkenazi populations. Although the individual subclades of the Ashkenazi populations are poorly supported statistically, their overall structure is logical. The French, Russian, UK and Ukrainian Jews are on the outer edge of this group while German Jews are at the base of other East Europeans groups. The Emigrant Europe populations are scattered throughout the European Ashkenazi. The large composite sample of Jewish populations ('Combined') is on the edge of the Ashkenazi clade, showing the substantial Ashkenazi composition of this group. Bulgarian Jews appear beyond all other European groups.

The North African Jews do not comprise a coherent group, but the contiguous geographic pairs Tunisia-Libya and Algeria-Morocco do appear, with the proximity of the latter to the European clade. Egyptian Jews form a nearly independent branch from which Iran, Iraq and Georgia appear. Turkey, like Egypt, is a nearly independent clade including Libya and Tunisia. Indian and Yemeni Jews are notable in their greatest divergence from the other Jewish populations.

Principal components analysis

In an effort to show structure within HLA haplotype frequencies across the Jewish populations, PCA was applied to the haplotype frequencies of 22 national samples of Jewish populations. PC1 accounted for 26.0% of the total variance, PC2 for 12.3%, and the first five PCs for 64.0%. PC1 separated the 12 European-origin Jewish populations from the other 10 Jewish samples, with a sizeable gap between the two groups (Figure 4). Across the axis of PC1 the Hungarian and Romanian Jews are outliers within the European Jewish samples. The sample of French-origin Jews while near the European group fell closest to the non-European-origin populations. Further population-specific HLA haplotype variability is distinguished by additional PCs. On PC2, the non-European populations showed the greatest divergence of Georgian and Libyan Jews at opposite extremes (Figure 4). Principal components PC3, PC4 and PC5 strongly distinguished the Jewish population samples of Yemen, India and Iran, respectively, pointing to the independent divergence of these gorups.

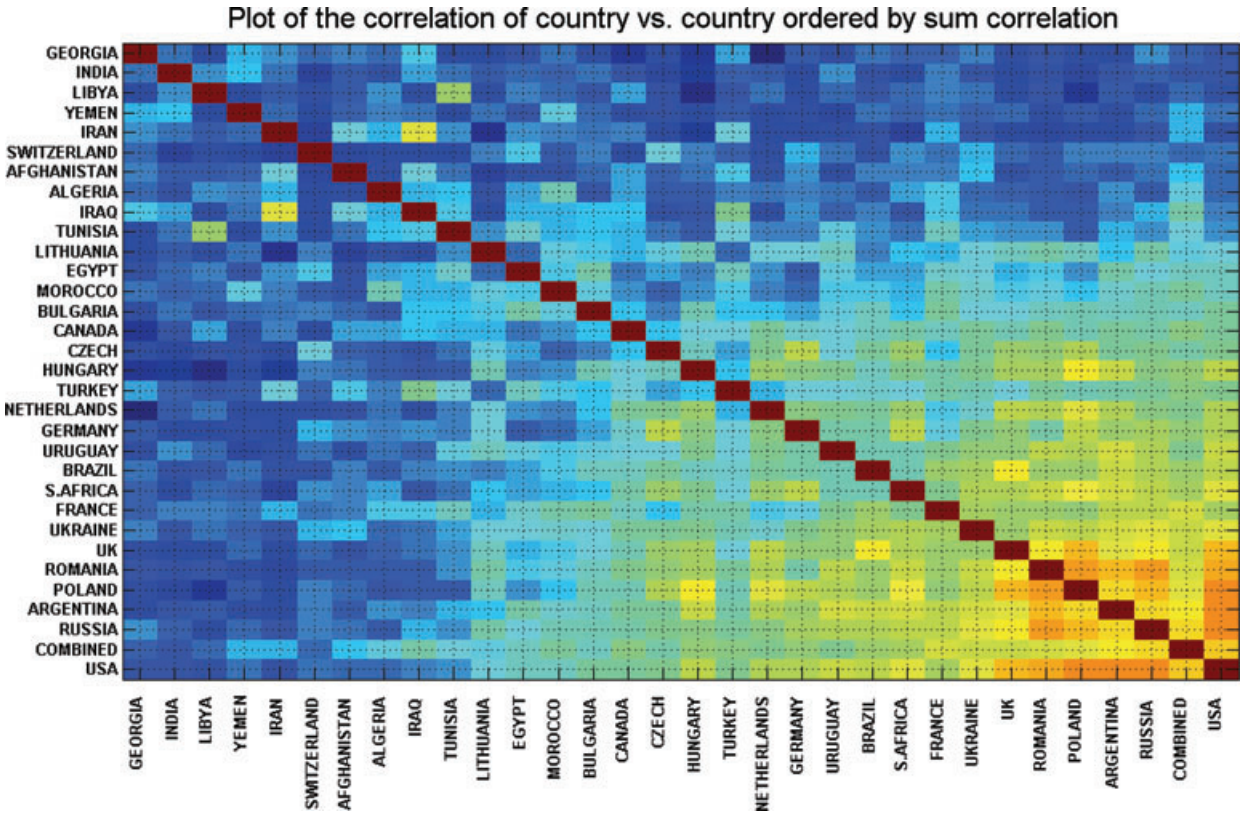

Figure 2 Plot of country-by-country Spearman rank correlation of haplotype frequencies ordered by the sum of the correlations ranging from high correlations (red) to low correlations (dark blue).

Simultaneous classification of populations and haplotypes

Figure 5 displays the eight clusters of haplotypes shown by simultaneous classification of populations and haplotype frequencies using the CLUTO software. In order to reduce sampling effects, only populations with at least 200 sampled haplotypes were used. The 23 populations in this analysis were comprised of 12 European-origin countries, and 10 other countries including origins in North Africa, the Middle East and India, as well as the agglomerated category 'Combined'. The first four (best delineated) groups (clusters 0–4) uniquely identified Jews from Morocco, Yemen, India and Georgia, respectively. These four groups were defined by 4–16 HLA haplotypes. Cluster 5 includes Libya and Tunisia and cluster six Iran and Iraq. In each of these clusters the pair of clustered countries shows some degree of sharing. Cluster six with its 21 haplotypes includes the several populations of European origin. Several other observations deserve mention. Turkey with haplotypes distributed across several other clusters did not fall with any single cluster. Cluster 7 includes both Egypt and Canada.

Table 4 summarizes the statistical support for the CLUTO analysis. The within-cluster statistic ISIM shows high within group concordance for each of the clusters. In contrast, the among cluster-statistic ESIM indicates varying degrees of support for each of the region-specific clusters, with the Morocco and Georgia showing at least modest support for the unique characteristic of their clustered haplotypes relative to all other haplotypes. The broad sharing seen for the sample of India-origin Jews, the Egyptian Jews and the Tunisia/Libya Jews seen in clusters 2, 4 and 5, do not show statistical support for the uniqueness of these clusters. For each of these three instances it can be seen that additional haplotypes are present in these clusters, rendering the cluster-defining haplotypes only part of the story for each cluster.

Besides the strong discrimination of the non-European source areas, a number of other results deserve mention (Figure 5). The Jews from France are part of the North African clade, and are especially close to the Moroccan Jews. The remaining European clade is then seen to be essentially Ashkenazi. Turkey bears haplotype affinities with its geographic neighbors, including Iran, Iraq and Georgia as well as others. Haplotypes of the category 'Combined' were seen to occur across all clusters, as is expected for this large sample consisting of contributions from all Jewish samples of undetermined national origin. The two most frequent haplotypes present in the Northern European host populations, 01∼08∼03 and 02∼44∼04, were each present in the Ashkenazi. The most

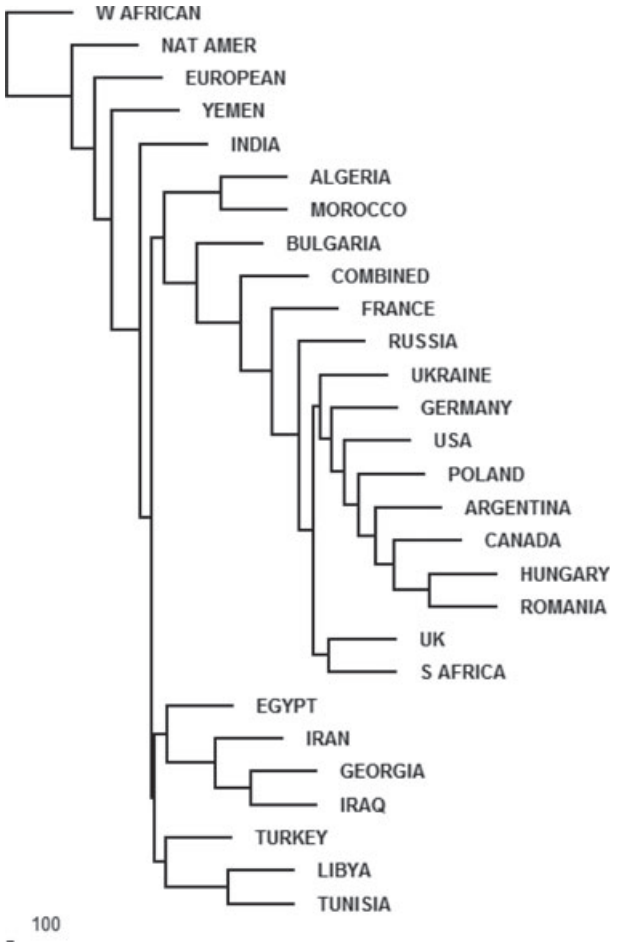

Figure 3 Neighbor-joining diagram of 25 Jewish populations defined by country of origin based on Nei's genetic distance of HLA A∼B∼DR haplotype frequencies. Samples from West Africans ($2n = 165{,}412$), Native Americans ($2n = 12{,}664$) and Europeans ($2n = 824{,}292$) are included as out-groups. The figure is based on the consensus tree derived from 1000 replicates. The scale shows the length of 100 replicates or 10%. For comparison, the terminal stems of all of the populations have a length of 1000 or all replicates.

common haplotype in the Ukraine, 02∼15∼06, also fell into the Georgian cluster, its Black Sea neighbor. Two haplotypes previously recognized as pan-Jewish each fell into the large European cluster (#7). The haplotype 26∼38∼04 was consistently present across the Ashkenazi populations, and was also observed in France and Morocco, but nowhere else. The haplotype 24∼35∼11 was present in 22 of the national samples, and absent only in Egypt.

Shared haplotypes among Jewish populations

By examining patterns of haplotype sharing across the Jewish diaspora (from Table S1 and Figure 5), it is possible to construct a list of candidate haplotypes largely common to Jews as a whole and to the Ashkenazi in particular (Table 5). Fourteen haplotypes are seen across many or most of the

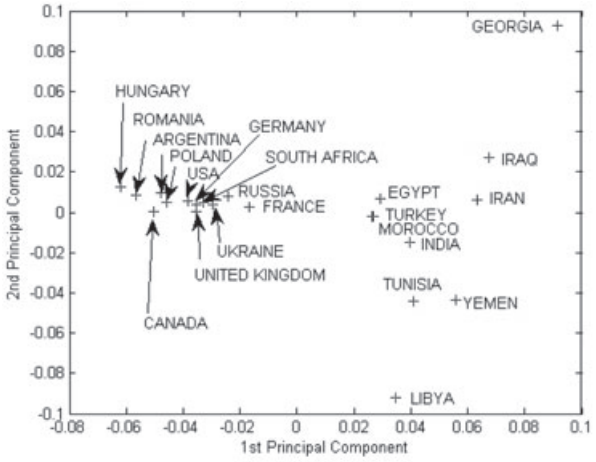

Figure 4 First two principal components of A∼B∼DRB1 haplotype variation in 22 country-of-origin Jewish populations.

population samples with 9 unique to Jewish, while 10 haplotypes appear to be largely unique or at least most prominent in Ashkenazi Jews. The previously mentioned haplotype 26∼38∼04 was found in the Ashkenazi, as well as many, but not all non-European population samples. Finally, three haplotypes were observed across most diaspora samples except for the Ashkenazi, of which four are familiar in non-Jewish populations. The construction of these haplotype groups (the presence or not of particular haplotypes across populations) was somewhat arbitrary in that additional haplotypes might be included or lost when stringency criteria were modified. Nonetheless, these categories are of value in identifying the common HLA characteristics of Jews. Another way of identifying Jewish haplotypes is to distinguish those haplotypes most responsible for population discrimination. The discriminatory haplotype groups (see clusters 0–6, Figure 5) are most informative for this purpose. We point out the presence of one haplotype of host European origin, 0101∼0801∼0301 in cluster 7 of European Jews. At low two-digit resolution, the haplotype 02∼44∼04 is a second potential candidate of European origin. While the high resolution A*0201 and B*4402 components of this haplotype match the common European haplotype, for most Jewish samples of this haplotype the DRB1*0402 allele is present and common in the Middle East, but quite rare in Europe. Another indicator of Eastern Mediterranean origins is the allele DRB1*0102, that is widespread across diaspora populations.

Discussion

We have used genetic data from the HLA complex, a single genetic system constituting 0.1% of the genome located at 6p21.3, and playing, within the genome as a whole, a central role in disease predisposition. Even in these early days of genome-wide assessments of disease predisposition a primary influence of HLA variation has been shown for metastatic

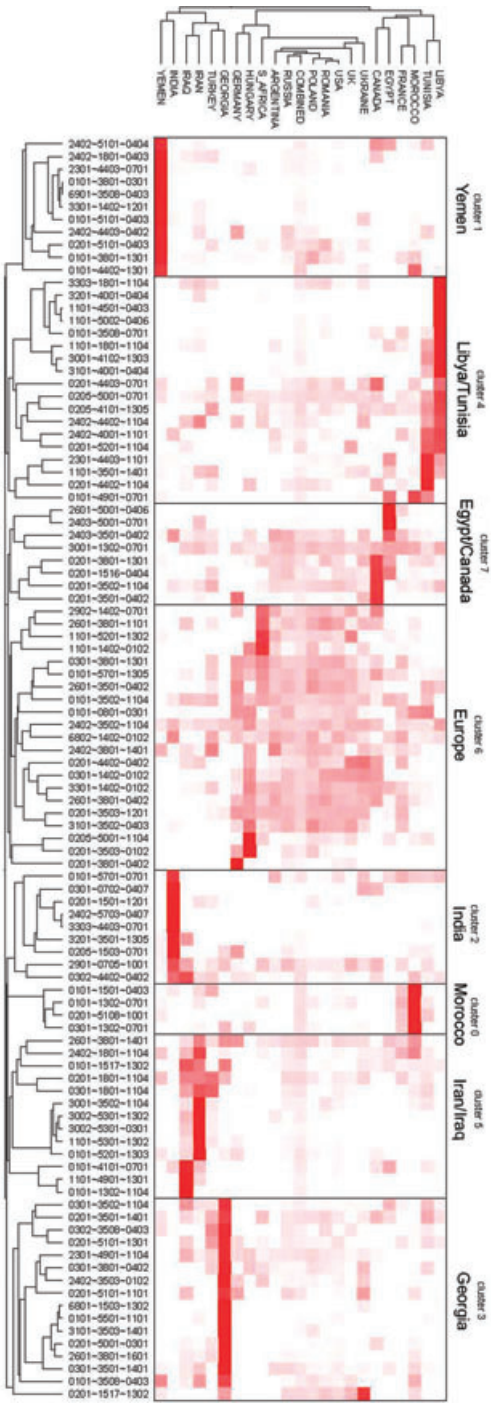

Figure 5 Clustering of the top 100 A~B~DR haplotypes on 22 selected Jewish populations defined by country of origin.

disease (39), autoimmune conditions (40, 41), and infectious disease (42), attesting to the functional importance of genetic variation in this region. The MHC may also stand out in terms of population divergence: In a study of population structure within the Japanese, a genome-wide screen on 140,000 single-nucleotide polymorphisms (SNPs) in 7003 individuals showed the HLA region to have the most informative markers of population differentiation (43).

The enormous population-level diversity of HLA haplotypes, with typically from 100s to 1000s of distinct haplotypes in a population (44), underlines the potential value of this single complex as a marker of population differentiation. A theoretical mechanism for this MHC characteristic of rapid population divergence has been offered (45), arguing that the driving force of heterozygote advantage in the MHC supplies a cover for linked deleterious recessives in the region. This operates to favor heterozygotes, even as population size and hence heterozygosity decrease, thus tending to increase population diversification. This present work shows in new detail and extent the ability of HLA haplotype variation to open a high-resolution window on stasis and divergence in the course of an elaborate population expansion and diversification event.

Jewish peoples emigrated from their ancestral home in the Eastern Mediterranean beginning over two and a half millennia ago, eventually establishing Jewish communities in regions across the globe. We examined 31 of these groups defined by contemporary country-of-origin information, and then studied their similarities and differences on the basis of HLA haplotypes. These original emigrants did not settle in unoccupied lands, but rather established new lives in regions already settled with previous residents or host populations. Our results show clear genetic differences among the Jewish populations and differing degrees of similarity among them. While a combination of population genetics forces must have contributed to the observed population divergence, a working hypothesis points to the important influence of admixture with host populations. Owing to the more or less distinct HLA compositions among the host populations, following admixture, this force would have enlarged genetic distances among members of the Jewish diaspora. Further insight on the validity of the admixture hypothesis will require concomitant studies of individual Jewish and host populations.

Evidence for genetic continuity among many diaspora populations was also identified. We have listed HLA haplotypes that are shared to a greater or lesser extent among Jewish populations generally, and the more recently diverged Ashkenazi groups in particular (Table 5). These haplotypes identify common genetic threads still present among Jews, and reflect the Middle Eastern source populations from which many Jewish people originate.

Integration of history and population genetics

It was possible to show similarities and differences among Jewish populations through a variety of methods for displaying relative population affinities. The neighbor joining, principal components, CLUTO and sum correlation analyses each tease out distinct aspects of Jewish population differentiation.

The highest sum correlations observed – among the European Jewish populations – are because of the fact that many

Table 4 Internal (ISIM) and External (ESIM) statistical support for the clusters depicted in Figure 5, including the region(s) most closely associated with each cluster

			ISIM			ESIM		
Cluster	Region(s)	Size	Mean	SD	*P*	Mean	SD	*P*
0	Morocco	4	0.950	0.011	<0.001	0.138	0.013	<0.01
1	Yemen	11	0.897	0.048	<0.001	0.085	0.042	ns
2	India	9	0.811	0.070	<0.001	0.115	0.089	ns
3	Georgia	16	0.814	0.100	<0.001	0.129	0.066	<0.10
4	Tunisia/Libya	18	0.653	0.143	<0.001	0.124	0.072	ns
5	Iran/Iraq	13	0.637	0.090	<0.001	0.148	0.085	ns
6	Europe	21	0.619	0.117	<0.001	0.169	0.074	ns
7	Egypt/Canada	8	0.657	0.088	<0.001	0.194	0.102	ns

Table 5 List of possible HLA A∼B∼DRB1 haplotypes in diaspora founder populations compiled from Table S1 and Figure 5

	Pan-Jewish	Ashkenazi	non-Ashkenazi
1	0101∼5701∼1305	2601∼3801∼0402	1101∼3501∼1401
2	3301∼1402∼0102	6802∼1402∼0102	2402∼1801∼1104[a]
3	0101∼3502∼1104	2601∼3501∼0402	0302∼4402∼0402
4	2402∼3502∼1104	2902∼1402∼0701	0101∼1517∼1302[a]
5	0205∼5001∼0701[a]	0301∼1402∼0102	0301∼1801∼1104[a]
6	2601∼3801∼1401	0301∼3801∼1301	2407∼3508∼1601
7	3001∼1302∼0701[a]	0201∼3503∼1201	0301∼3502∼1104
8	2403∼3501∼0402	3101∼3502∼0403	
9	2402∼3801∼1401	1101∼1402∼0102	
10	0101∼5701∼0701[a]	1101∼5201∼1302	
11	2901∼0705∼1001[a]		
12	0201∼1801∼1104		
13	0201∼3501∼1401		
14	0201∼3502∼1104		

[a]Widespread haplotypes also present in non-Jewish populations.

European Jewish samples are present in the study and that many of these groups (especially the Ashkenazi) are highly similar (Figure 2). The greatest degree of similarity overall (shared by Combined and USA) is apparently because of the mixed – albeit Ashkenazi dominated – composition of each. The axes of independently explained variance in haplotype frequencies showed from the PCA (Figure 4) show two fundamental features of Jewish population divergence: first a distinct range of the numerically dominate European Jewish populations, and second, the diverse composition of the non-European Jewish populations underscored by the extreme placement of individual populations with the higher order principal components. The neighbor-joining analysis (Figure 3) also affirms those basic patterns, as well as showing the historical and geographic logic underlying Jewish population differentiation. For example, we see a central derived clade of Ashkenazi dominate groups with UK, Russia and France on the edges, followed by the mixed Combined and then the most divergent European group, the Bulgarian Jews. Extending beyond the European clade, geographically proximate subgroups are seen, including Iran/Iraq/Georgia, and the North African pairs Algeria/Morocco and Libya/Tunisia. The proximity of Algeria/Tunisia to the European clade might be a reflection of population movements during the centuries of Moorish Iberia and from the Jewish population exiled from Iberia during the 14th and 15th centuries to regions in North Africa. Two Jewish haplotypes 2601∼3801∼0402 and 6802∼1402∼0102 are found in Ashkenazi, North Africa, Turkey and Bulgaria, each of which absorbed Jews from the Iberian exile. Yemeni and Indian Jews appear as quite distinct, lying just on the Jewish side of the out-group control samples.

The clustering algorithm CLUTO, integrating information from both population and haplotype assignments, sheds further light. The Moroccan Jews are identifiable by four haplotypes – each of which is shared by the French Jews. The unique clusters of haplotypes associated with Yemen, India and Georgia attest to their distinctness. The broad haplotype affinities of Libya/Tunisia and Iran/Iraq further underscore the commonalities of these pairs (clusters 4 and 5 respectively). The poorly differentiated members of the European Jewish group (cluster 6) with its 21 associated haplotypes fall out next. The last cluster, 7, contains Egypt and Canada, an unanticipated association. This is because of the sharing of three haplotypes, while the other five of the eight haplotypes in the cluster are entirely unshared. In a CLUTO analysis of the European and Emigrant-European Jewish populations alone (data not shown), Canada – at cluster 0 – is a distinctive outlier from the other European samples. This suggests that some fraction of Canadian Jews may have emigrated from the Jewish population of Egypt. Finally, the one non-European population not part of a cluster is Turkey which has haplotypes spread across the full range of the other populations. This must reflect the diverse affinities and origins of the Turkey Jews.

A great deal of scholarly work is available on Jewish history, documenting many details of the origins, growth and decline of diaspora populations. We mention some important historical touchstones useful for understanding the population genetic structure of Jews (3–5). Jewish founders derive from various population sources of the Eastern Mediterranean region. The intimate economic and geographic association of

Jewish peoples with their host populations typically stretching over many generations appears to have resulted in admixture, a tendency inherent in the social and cultural circumstances of the population encounters. Contrary to the notion of a narrowly and biologically defined 'chosen people' and adding to the idea that the Jewish people are in fact an amalgam of many population sources, is the story that Abraham himself, considered the early father of Judaism (and monotheism generally), came from Babylonia (modern day Iraq). Several population relationships among Jews (Figures 2–4) can be interpreted in the light of historical events. First, outlier populations, including those of Yemen and India (haplotype clusters, Figure 5), may reflect their distance from the Eastern Mediterranean heartland in terms of host population divergence, the long period of isolation in each case, and the possibility of direct foreign ancestry through conversion. The genetic similarity of the Iranian and Iraqi populations (both are members of cluster 6, Figure 5) may reflect common threads going back to the Babylonian exile. Jewish populations from Georgia, that are also members of the Iran–Iraq sub-clade (Figure 3), may reflect contributions from that same ancient strand of early Jewish migrants. The remaining non-Ashkenazi populations are, except for French Jews, all North African, and display a distinct set of relationships. The affinities of Moroccan and French Jewish populations (Figure 5) are likely because of the movement of Sephardic Jews following the 14th and 15th century expulsions from Iberia. The position of the sample of Egyptian Jews with its rather unique haplotypic signature (cluster 7, Figure 5) may be a consequence of both the ancient implant of Jews into Egypt and the long duration of that contact.

Genetic differences found among the Ashkenazi populations can also be interpreted in the light of available history (Figures 2 and 5). The more subtle distinctions among the Ashkenazi are a consequence of fewer generations of separation between the Jews that have inhabited the modern nation states of Central Europe, and are likely accentuated by the more uniform genetic background of the Ashkenazi host populations. The proximity of Russian and Combined (Figures 2 and 5) may reflect the considerable contribution of Russian–Jewish immigrants to the Israeli Combined group, and similarly for the US–Polish clades (Figures 2, 3 and 5). The similarities of members of the Polish and Romanian clades (along with the recent US and Argentinean offshoots) may hint at the population expansions of the Ashkenazi across the Polish-Lithuanian Commonwealth during the 16th and 17th centuries (46). It is possible that the ordered relationship in the cladogram based on Nei genetic distances – Germany, Poland, then Hungary/Romania (Figure 3) – reflects the historical movement of the Ashkenazi across Europe. The Jewish populations of France, Ukraine and the UK falling at the Ashkenazi periphery (Figure 3) may reflect the combination of diverse origins (Sephardic, Ashkenazi and Near-East Jewish sources) and distinct host populations.

The study of Y chromosome variation in contemporary Iberian males (14) mentioned above showed 20% average prevalence of Jewish Y haplotypes across Portuguese and Spanish Provinces reflecting Jewish – host population admixture accruing over the 600 years of Moorish occupation and its aftermath. The dispersed 'Sephardic' population was the recipient of unknown fractions of Moorish and European genetic influence over this same period. We posit that a concomitant population genetic theme of all diaspora Jewish populations is the sharing of genetic material with host populations. This phenomenon likely supplies the strongest explanation for the observed rapid divergence of diaspora populations.

The great majority of genetic variation is shared across all people (47). The identification of that fraction of markers alone or in combination capable of reflecting an individual's ancestral population inheritance should not be confused with the fact that many individuals in an ethnic group may not share those threads of common ancestral genetic background, but are nonetheless full-fledged members of that community. We make that point here with the demonstration of the considerable genetic divergence of Jewish people: All are culturally Jewish, while only a subset bear specific indicators of a genealogical history originating in ancient Palestine.

Practical implications for bone marrow registries

One direct application of these results is to the design and functioning of bone marrow registries for allogeneic stem cell transplantation. Extensive studies repeatedly show that complete HLA matching of donor and recipient favors successful transplantation. Haplotype variation of HLA is both extensive and population specific. For these reasons donor registries must include large numbers of individuals and be population specific.

The Jewish population of modern Israel is a consequence of a process which brought Jews from the worldwide diaspora to Israel. The data presented in this study show that the Jewish people differ genetically among themselves according to their ethnicity, but also, in many instances, do share some common characteristics in the HLA complex. In this regard, it is evident that there is far greater similarity between the different Ashkenazi populations than between the non-Ashkenazi Jewish populations (Figures 2–5), while at the same time both Ashkenazi and non-Ashkenazi Jews differ from non-Jewish Caucasians. The distinct population clusters shown in Figure 5 illustrate the need for both Ashkenazi and specific non-Ashkenazi groups in donor registries in order to maintain a matchable reservoir of unrelated hematopoietic stem cell donors for Jewish patients.

Comparison to genome-wide studies of Jewish population differences

Two works have appeared describing genome-wide surveys of genetic relationships among Jewish populations (48, 49).

Each of these note the considerable degree of genetic correspondence of the major Jewish clades to populations of the Levant, especially the Druze and Palestinians. Many of the results and conclusions of these studies correspond with the findings we describe here, including the identification of Ashkenazi, Middle Eastern, and Sephardic-derived strands of the historical Diaspora and the recognizable threads of Jewish genetics leading back to the Levant. Each study recognizes the existence of Jewish populations with few or no links to the rest, such as the Indian Jews. In addition, the studies recognized admixture with host populations as a force driving Jewish population divergence.

It should be kept in mind that the occurrence of selection following admixture may differentially favor genes from one or the other of parental populations (50). Thus, HLA proportions in an admixed group such as a diaspora Jewish population may not exhibit exogamous HLA frequencies at the same rate as (unselected) non-HLA genetic markers. This same phenomenon may have operated with the CCR5 D32 variant that appears to have moved into Jewish populations to protect against small pox (51). Similarly, our analysis offers possible specific causative agents (identified HLA specificities) worthy of pursuit in understanding the pathogen-driven selection environments experienced by both host and the diaspora Jewish populations.

Conclusion

The differentiation of the populations of the Jewish diaspora recorded here appears to be because of two principal sets of causes. First the ability to construct a list of identifiable Ashkenazi and non-Ashkenazi haplotypes (Table 5) reflects the differential survival of founder haplotypes among Jewish populations from the original emigrants representing the ancient Israeli homeland across Jewish ethnic groups indicating a common background. Contributing forces include the historical contingencies favoring one haplotype over another as populations expanded and split. The survival and increase of some haplotypes must have been influenced by differences in the demographic successes of the various populations. Natural and cultural selection acting on some haplotypes may also have played a role. But as the HLA evidence suggests, a second factor – admixture with host populations – may have been an important force driving the differentiation of Jewish populations over a relatively brief number of generations. We posit that this is likely to have been a primary cause of the observed divergence. The variation in haplotypes among populations (both the presence and frequency of differences between populations) responsible for the divergence (Figures 1, 3, 4 and 5) did not originate in ancient Palestine, but rather were picked up during the population residences of Jews throughout the diaspora. This argument will require future studies involving specific comparisons of Jewish and host populations for confirmation.

Acknowledgments

Dedication: This publication marks life-time achievements of the coauthor Dr Chaim Brautbar in the establishment of the Hadassah Unrelated Hematopoietic Stem Cell Donor Recruitment Program and Registry and his contributions to the study of HLA Jewish population anthropology.

We thank Carolyn Hoppe and Janet Poff for their valuable comments on the manuscript and Irina Haviv, senior programmer, Department of Information Systems, Hadassah University Hospital, Jerusalem for her professional support. We thank Michael Steinbach and Vipin Kumar from the University of Minnesota Computer Science Department for their technical expertise on the CLUTO and sum correlation methods. We also thank the four reviewers for their perceptive and useful comments for improving the manuscript. Supported in part by Office of Naval Research contract N00014-08-058.

References

1. Meyer D, Single R, Mack SJ, Erlich HA, Thomson G. Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genetics* 2009: **173**: 2121–42.
2. Cohen S. *The Beginnings of Jewishness – Boundaries, Varieties, Uncertainties. Berkeley*: *University of California Press*, *1999*.
3. Ben-Sasson HH. *A History of the Jewish people*. *London*: *Weidenfeld and Nicolson*, *1977*.
4. Roth C. *A History of the Jews: From Earliest Times Through the Six Day War. New York*: *Schocken Books*, *1971*.
5. Biale D, ed. *Cultures of the Jews: A New History. New York*: *Schocken Books*, *2002*.
6. Ben-Naeh Y. Blond, tall, with honey-colored eyes: Jewish ownership of slaves in the Ottoman Empire. *Jew His* 2006: **20**: 73–90.
7. Sabar S. Childbirth and Magic: Jewish Folklore and Material Culture. In: Biale D, ed. *Childbirth and Magic: Jewish Folklore and Material Culture. New York*: *Schocken Books*, *2002*, *369–419*.
8. Meyer D, Single R, Mack SJ et al. 13th IHWS anthropology/human genetic diversity joint report. Single locus polymorphism of classical HLA genes. In: Hansen JA, ed. *Immunobiology of the Human MHC. Seattle*: *International Histocompatibility Working Group Press*, *2006*, *653–704*.
9. Thomas MG, Parfitt T, Weiss DA et al. Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba – the 'Black Jews of Southern Africa'. *Am J Human Genet* 2000: **66**: 674–86.
10. Thomas MG, Skorecki K, Ben Ami H, Parfitt T, Bradman N, Goldstein DB. Origins of old testament priests. *Nature* 1998: **394**: 138–140.
11. Behar DM, Thomas MG, Skorecki K et al. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *Am J Hum Genet* 2003: **114**: 354–65.
12. Behar DM, Hammer MF, Garrigan D et al. MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur J Hum Genet* 2004: **12**: 335–364.

13. Feder J, Blech I, Ovadia O et al. Differences in mtDNA haplotype distribution among 3 Jewsish populations alter susceptibility to T2DM complications. *BMC Genomics* 2008: **9**: 198.
14. Adams SM, Bosch E, Balaresque PL et al. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet* 2008: **83**: 1–12.
15. Klitz W, Gragert L, Maiers M et al. Four-locus high-resolution HLA typing in a sample of Mexican Americans. *Tissue Antigens* 2009: **74**: 508–13.
16. Tian C, Plenge RM, Ransom M et al. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet* 2008: **4**: e4.
17. Price AL, Butler J, Patterson N et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 2008: **4**: e236.
18. Need AC, Kasperaviciute D, Cirulli ET, Goldstein DB. A genome-wide genetic signature of Jewish ancestry perfectly separates individuals with and without full Jewish ancestry in a large random sample of European Americans. *Genome Biol* 2009: **10**: R7.
19. Bonné-Tamir B, Bodmer JG, Bodmer WF et al. HLA polymorphism in Israel 9, An overall comparative analysis. *Tissue Antigens* 1978: **11**: 235–250.
20. Tabatabai H, Mohammad K, Mohagheghpour N. HLA antigens in two Iranian populations: the Armenians and the Jews. *Tissue Antigens* 1978: **12**: 309–314.
21. Cohen T, Levene C, Yodfat Y et al. Genetic studies of Cochin Jews in Israel: 1. Population data, blood groups, isoenzymes and HLA determinants. *Am J Med Genet* 1980: **6**: 61–73.
22. Levene C, Steinberg AG, Friedlander Y, Brautbar C, Cohen T. Genetic polymorphisms among Bukharan and Gerogian Jews in Israel. *Am J Med Genet* 1984: **19**: 623–41.
23. Brautbar C, Battat S, Sherman L, Benhamu R, Cohen O. HLA antigens in Israeli Ashkenazi and non-Ashkenazi Jews. In: Aizawa M, ed. *HLA in Asia-Oceania, Proceedings of the third Asia Oceania Histocompatibility Workshop Conference. Sapporo*: *Hokkaido:University Press*, *1986*, *324–7*.
24. Roitberg-Tambur A, Friedmann A, Witt CS et al. HLA polymorphism in Moroccan Jewry. *Hum Immunol* 1994: **40**: 61–7.
25. Cox ST, Marsh SG, Scott I et al. HLA-A, -B, -C polymorphism in a UK Ashkenazi Jewish potential bone marrow donor population. *Tissue Antigens* 1999: **53**: 41–50.
26. Martinez-Laso J, Gazit E, Gomez-Casado E et al. HLA DR and DQ polymorphism in Ashkenazi and non-Ashkenazi Jews: comparison with other Mediterraneans. *Tissue Antigens* 1996: **47**: 63–71.
27. Roitberg-Tambur A, Witt CS, Friedmann A et al. Comparative analysis of HLA polymorphism at the serologic and molecular level in Moroccan and Ashkenazi Jews. *Tissue Antigens* 1995: **46**: 104–110.
28. Fort M, de Stefano GF, Cambon-Thomsen A et al. HLA class II allele and haplotype frequencies in Ethiopian Amhara and Oromo populations. *Tissue Antigens* 1997: **51**: 327–336.
29. Klitz W, Gragert L, Maiers M. Re-creation of the genetic composition of a founder population. *Human Genetics* 2008: **124**: 417–21.
30. Kollman C, Maiers M, Gragert L et al. Estimation of HLA-A, -B, -DRB1 haplotype frequencies using mixed resolution data from a National Registry with selective retyping of volunteers. *Hum Immunol* 2007: **68**: 950–8.
31. Hurley CK, Setterholm M, Lau M et al. Hematopoietic stem cell donor registry strategies for assigning search determinants and matching relationships. *Bone Marrow Transplant* 2004: **33**: 443.
32. Gourraud P, Génin E, Cambon-Thomsen A. Handling missing values in population data: consequences for maximum likelihood estimation of haplotype frequencies. *Eur J Hum Genet* 2004: **12**: 805–12.
33. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992: **48**: 361–72.
34. Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for un phased diploid genotype data. *Am J Hum Genet* 2000: **67**: 947–59.
35. Nei M. Genetic distances between populations. *Am Nat* 1972: **106**: 283–92.
36. Felsenstein J. PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* 1989: **5**: 164–6.
37. Novembre J, Stephens M. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet* 2008: **40**: 646–649.
38. Karypis G. CLUTO: a clustering toolkit. Technical Report 02-017, 2002. University of Minnesota.
39. Tse K-P, Su W-H, Chang K-P et al. Genome-wide association study reveal multiple nasopharyngeal carcinoma-associated loci within the HLA region at chromosome 6p21.3. *Am Soc Hum Genet* 2009: **85**: 194–203.
40. Zuvuch RL, McCauley JL, Pricak-Vance MA, Haines JL. Genetics and pathogenesis of multiple sclerosis. *Semin Immunol* 2009.
41. Feng B-J, Sun L-D, Soltani-Arabshahi R et al. Multiple loci within the major histocompatibility complex confer risk of psoriasis. *PLoS Genet* 2008: **5**: e1000606.
42. Limou S, LeClerc S, Coulonges C et al. Genomewide association study of an AIDS nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis* 2009: **199**: 419–26.
43. Yamaguchi-Kabata Y, Nakazono K, Takahashi A et al. Japanese population structure based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet* 2008: **83**: 445–56.
44. Maiers M, Gragert L, Klitz W. High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol* 2007: **68**: 779–88.
45. Van-Oosterhout C. A new theory of MHC evolution: beyond selection on the immune genes. *Proc R Soc B* 2009: **276**: 657–65.
46. Rosman M. Innovative tradition: Jewish culture in the Polish-Lithuanian Commonwealth. In: Biale D, ed. *Cultures of the Jews: A New History. New York*: *Schocken Books*, *2002*, *519–72*.

47. Lewontin RC. *The Genetic Basis of Evolutionary Change. New York*: *Columbia University Press*, *1974*.
48. Behar DM, Yunusbayev B, Metspalu M et al. The genome-wide structure of the Jewish people. *Nature* 2010: **466**: 238–42.
49. Atzmon G, Hao L, Pe'er I et al. Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am J Hum Genet* 2010: **86**: 850–859.
50. Baus A, Tang H, Zhu X et al. Genome-wide distribution of ancestry in Mexican Americans. *Hum Genet* 2008: **124**: 207–14.
51. Klitz W, Brautbar C, Scyhito AM, Barcellos LF, Oksenberg JR. Evolution of the CCR5*D32 mutant based on haplotype variation in Jewish and European population samples. *Human Immunol* 2001: **62**: 530–8.

Supporting Information

The following supporting information is available for this article:

Table S1. The 100 most common A∼B∼DR haplotypes in 30 Jewish populations.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

